**Biol251 - STATISTICS ASSIGNMENT 2021 (out of 120)**

**Question 1 – (20 marks)**

1. Ten replicate bivalves were placed into containers that contained one of 5 different salinities (50 animals in total) in an experiment to evaluate survival in estuaries. Half of the animals in each treatment were placed in containers with 25C water and half were in containers with 15C water. Shell growth was measured after 5 months. The 2-factor analysis of variance produced a significant interaction term. Describe how you would undertake multiple comparisons and what Tukeys critical value you would use for the test. (4 marks)
2. How is the F ratio calculated? (2 marks)
3. What is the SS and how is it related to the MS? (2 marks)
4. In a one factor ANOVA how is the MS related to the variance (2 marks)?
5. In the example in Question 1 (above), calculate the proportion of the variation in the model explained by a) heat, b) drought and c) the interaction between heat and drought? When these 3 values are added together what is the meaning of the value that you obtain (6 marks)
6. In a two factor analysis of covariance, what is the hypothesis that the F ratio of the interaction term tests? (4 marks)

**Question 2 – (24 marks)**

An urban ecologist was interested in the stresses of urban living on the growth of shrubs. An experiment was undertaken investigating the effect of drought and heat stress on growth rates of a native shrub. Thirty two shrubs were divided into 2 heat wave treatments: 8 at standard temperature and 8 at heat wave conditions for each of 2 levels of drought; water every 8 days and daily watering. After a month shrubs were harvested and weighed and weight recorded. The results are:

Heat wave, drought

15, 16, 13, 19, 16, 13, 19, 20, g

Heat wave, no drought

32, 31, 24, 28, 21, 26, 19, 18 g

Normal temp, drought

21, 24, 21, 20, 29, 34, 26, 15 g

Normal temp, no drought

32, 30, 17, 22, 23, 31, 27, 30 g

Given this experiment:

1. What are the hypotheses being tested with this experiment? (3 marks)
2. What is the word equation for the model being entered into the statistics program? (1 mark)
3. Undertake the analysis using JMP or similar statistics program and present a printout of the results and any other workings. This includes appropriate assumptions and multiple comparisons (10 marks)
4. Present the graph that is most appropriate in showing this interpretation with legend (6 marks)
5. What issues might arise in allocating plants to particular treatments at the beginning of the experiment. How would you overcome this (4 marks)

**Question 3 (16 marks)**

Previous observations suggested that a crayfish that inhabits seagrass beds has a negative effect on survival of a range of polychaetes. The interaction between the presence of seagrass beds and the survival of a polychaete was assessed in 3 estuaries. In each estuary there were 2 habitat types that represented different habitat types: one within and one away from a seagrass bed. In these sites a single measure of survival was calculated based on the no. of surviving polychaetes from 100 translocated polychaetes. Crayfish abundance was assessed using transect surveys in each seagrass patch. The JMP analysis has been printed out for you together with a graph.

1. What is the total number of samples taken? How many replicates within each estuary x habitat type combination? (2 marks)
2. Which independent factors are categorical? Which are continuous? (3 marks)
3. State the hypotheses that are supported by the analysis? (4 marks)
4. Are the crayfish having a neutral, positive or negative influence? (2 marks)
5. Is estuary important? (2 marks)
6. Is the graph appropriate based on the results? What other things ought to be added to interpret the results graphically (3 marks)

**Response growth rate**

**Whole Model**

**Summary of Fit**

|  |  |
| --- | --- |
| RSquare | 0.749735 |
| RSquare Adj | 0.596795 |
| Root Mean Square Error | 1.577143 |
| Mean of Response | 4.883333 |
| Observations (or Sum Wgts) | 30 |

**Analysis of Variance**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** |
| --- | --- | --- | --- | --- |
| Model | 11 | 134.12881 | 12.1935 | 4.9022 |
| Error | 18 | 44.77286 | 2.4874 | **Prob > F** |
| C. Total | 29 | 178.90167 |  | 0.0015\* |

**Effect Tests**

| **Source** | **Nparm** | **DF** | **Sum of Squares** | **F Ratio** | **Prob > F** |  |
| --- | --- | --- | --- | --- | --- | --- |
| Crayfish abundance | 1 | 1 | 1.333217 | 0.5360 | 0.4735 |  |
| Estuary | 2 | 2 | 9.923256 | 1.9947 | 0.1650 |  |
| seagrass | 1 | 1 | 17.064878 | 6.8606 | 0.0174 |  |
| Estuary \* seagrass | 2 | 2 | 19.592561 | 3.9384 | 0.0381 |  |
| Estuary \* Crayfish abundance | 2 | 2 | 0.067872 | 0.0136 | 0.9865 |  |
| seagrass \* Crayfish abundance | 1 | 1 | 0.017986 | 0.0072 | 0.9332 |  |
| Estuary \* seagrass \* Crayfish abundance | 2 | 2 | 4.332775 | 0.8710 | 0.4355 |  |

**Question 4 (14 marks)**

A study was designed to investigate the reproductive success of noisy miners in natural and plantation woodlands. Four natural and four plantations were surveyed and 5 females within each area, each from different family groups were followed to determine how many young fledged over a 3 year period. As the species is a cooperative breeder, size of the group in each territory was measured to investigate how group size might influence fitness of each female.

1. What is the dependent variable? (1 marks)
2. What are the independent variables to be put into the model? (4 marks)
3. Which ones are
   1. fixed,
   2. random,
   3. nested,
   4. categorical
   5. continuous? (5 marks)
4. In analyzing the data, the ecologist found that the ‘size of group’ by habitat interaction term was significant. Describe what this means? (4 marks)

**Question 5 (24 marks)**

Take the data available on moodle called ‘Assignment Q5 sleepy lizards’ in a worksheet called ‘sleepy lizards’. This data measured the weight of sleepy lizards on 4 properties in South Australia. Sleepy lizards were assessed to determine whether their weight was influenced by grazing. Properties with high and low grazing regimes (2 of each) were each searched to find three colonies of lizards. Within each colony, lizards from 3 different status groups were caught and weighed and while their status within the colony was assessed, it was initially ignored in the analysis.

1. What is the word equation for the model described above (without status included)? (3 marks)
2. Which factors are nested? (2 marks)
3. Which factors are random? (2 marks)
4. How will the nested and random factors influence the analysis? (3 marks)
5. Run the analysis (without status included) and present the F table. Put an asterisk next to any F ratios that are incorrect (you don’t have to recalculate these!)[hint: are assumptions for test met?]. (10 marks)
6. A final column (ignored in your analysis but nevertheless present in the dataset!) identified the status of the lizards. If this factor was included in the analysis would it be fixed or random and would it be nested or orthogonal (4 marks)

**Question 6. (Extension question - 15 marks)**

A study investigated survival of coastal grass seedlings in areas of high and low salinity. As endomycorrhizal associations are found in all grass spec24+ies, it was hypothesized that increase mycorrhizal colonization would improve the chances of survival. Survival was measured after a month growing in one of two salinity treatments. Roots were harvested and assessed to determine the percentage colonization of arbuscular mycorrhizal fungi. The results are found in the worksheet ‘coastal grass’ which records whether the plant was alive (1) or dead (0) at the end of the experiment, the salinity experienced and the % colonization of roots. Bring the data into JMP. Change the column type for alive to be categorical.

1. Fit the appropriate model to the data. What is the dependent variable? Is it continuous or categorical? (2 marks)
2. When you put it in the Y column what does JMP say is the ‘personality’ it will use to analyse the data? (1 mark)
3. What is the word equation of the model? (2 marks)
4. What statistic is used in this kind of analysis to test the hypothesis? (2 mark)
5. This analysis tests whether the probability of being alive is influenced by your factors. What factors were found to influence the probability of being alive? Attach you JMP printout to show this. (4 marks)
6. What is the R2 (or U) value of this model. (2 marks)
7. What is an AIC? What is an AICc? What is the AICc in this model? (2 marks)

If you have done this correctly, then you have achieved a **binomial logistic regression**……. Well done.!

**Question 7. (7 marks)**

**For you quadrat size and number exercise in week 4;**

Identify the organism and habitat that you sampled.

Present your figure to **identify** the most appropriate quadrat size to use to measure the abundance of your organism in the habitat. Mark where your decision lies.

Present your figure to **identify** the number of quadrats needed to measure the abundance of your gastropod on the rocky shore. Mark where your decision lies.